Course One
Foundations of Data Science

Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☑ ~~Complete the PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.~~

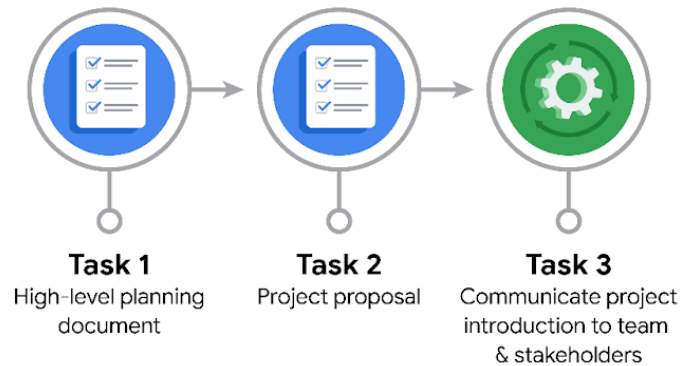- ☑ ~~Create a project proposal for the data team.~~

Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?

- How would you plan an analytics project?

- What steps would you take to translate a business question to an analytical solution?

- Why is actively managing data an important part of a data analytics team's responsibilities?

- What are some considerations you might need to be mindful of when reporting results?

Reference Guide

This project has three tasks; the following visual identifies how the stages of PACE are incorporated across those tasks.

Task 1
High-level planning document

Task 2
Project proposal

Task 3
Communicate project introduction to team & stakeholders

Data Project Questions & Considerations

PACE: Plan Stage

- Who is your audience for this project?

  - TikTok Data Team (technical)
  - Project Management Officer (non-technical)
  - Finance Lead, Americas (non-technical)
  - Operations Lead (non-technical)

- What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

  Build and deliver a reliable machine learning model to classify user interaction data as a claim or opinion. The impact will be a reduction in the backlog of user reports and improved efficiency in content moderation.

- What questions need to be asked or answered?

  - What are the specific features available in the dataset?
  - Are there any missing or incomplete data?
  - What type of regression model is most appropriate for this data?
  - What are the key metrics for model evaluation?
  - What are the expectations for the final deliverables and presentation format?
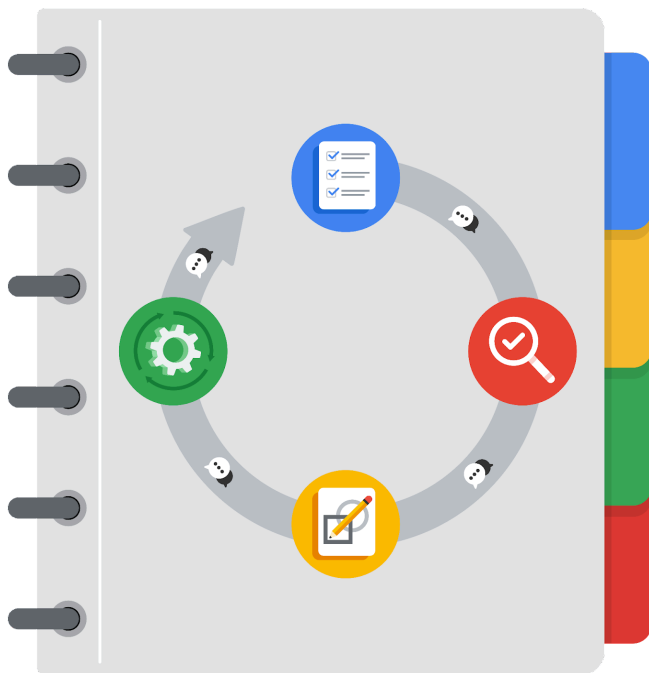
- What resources are required to complete this project?

  > - Access to the dataset
  > - Python environment setup for data analysis and model building
  > - Data visualization tools (e.g., Tableau)
  > - Collaboration tools (e.g., Slack, Zoom)

- What are the deliverables that will need to be created over the course of this project?

  > - Project proposal document
  > - Exploratory Data Analysis (EDA) report
  > - Hypothesis testing results
  > - Machine learning model and its evaluation metrics
  > - Visualizations for presentation
  > - Final report with insights and recommendations

THE PACE WORKFLOW

[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the Course 1 end-of-course portfolio project overview reading if you need more information about the tasks within the project.

Project tasks

Following are a group of tasks your company's data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following readings to help guide your selections and explanation: The PACE stages and Communicate objectives with a project proposal. You will later reorder these tasks within a project proposal.

1. **Evaluating the model:** **Construct** ▾

   Why did you select this stage for this task?

   This task involves assessing the performance of the machine learning model, which is a construction activity since it requires analysis of the model's outputs.

2. **Conduct hypothesis testing:** **Analyze** ▾ and **Construct** ▾

   Why did you select these stages for this task?

   Hypothesis testing involves analyzing the data to validate assumptions and then constructing insights based on the results of the tests.

3. **Begin exploring the data:** **Analyze** ▾

   Why did you select this stage for this task?

   Data exploration is the initial analysis to understand the dataset's structure, identify patterns, and detect any anomalies or missing values.

4. **Data exploration and cleaning:** **Analyze** ▾ and **Construct** ▾

   Why did you select these stages for this task?

   This task involves analyzing the data to identify issues and then constructing a cleaned version of the dataset for further analysis.

5. **Establish structure for project workflow (PACE):** **Plan** ▾

Why did you select this stage for this task?

This task involves analyzing the data to identify issues and then constructing a cleaned version of the dataset for further analysis.

6. **Communicate final insights with stakeholders:** Execute ▾

Why did you select this stage for this task?

This task involves executing the communication strategy to share the final results and insights with stakeholders.

7. **Compute descriptive statistics:** Analyze ▾

Why did you select this stage for this task?

Calculating descriptive statistics is part of the data analysis process to summarize and describe the main features of the dataset.

8. **Visualization building:** Construct ▾ and Execute ▾

Why did you select these stages for this task?

Creating visualizations involves constructing the visuals and then executing their presentation to stakeholders.

9. **Write a project proposal:** Plan ▾

Why did you select this stage for this task?

Drafting the project proposal is a planning activity that outlines the project's goals, tasks, and deliverables.

10. **Build a regression model:** Construct ▾ and Analyze ▾

Why did you select this stage for this task?

Constructing the model involves creating it, while analyzing involves evaluating its performance and making adjustments.

11. **Compile summary information about the data:** **Analyze** ▾

Why did you select this stage for this task?

> Summarizing the data is an analysis task to provide an overview of the dataset.

12. **Build machine learning model:** **Construct** ▾

Why did you select this stage for this task?

> **This involves the actual construction of the machine learning model using the cleaned and prepared data.**